





Introdução a Text Mining (TM)

Bem-vindo ao nosso estudo sobre abordagens de Text Mining - TM. Este estudo ajudará você a compreender melhor como estudar a disciplina de Desenvolvimento de Algoritmos de Inteligência Artificial. Neste módulo será abordado uma breve introdução à TM, seus objetivos e conceitos e como é a seleção de tipos de informações.

Conhecer os conceitos que serão apresentados aqui te ajudará a chegar ao sucesso do seu estudo. Vamos começar a estudar?

Introdução

Com o aumento do número de informações geradas pelos programas, aplicações, sistemas, aplicativos mobile e o uso contínuo da internet é possível perceber a necessidade de evolução da Recuperação de Informações. Text Mining ou Mineração de Texto utiliza técnicas de análise e extração de dados a partir de palavras, frases ou textos, faz parte do processo de Descoberta do Conhecimento. Mas você deve estar se perguntando, o que Text Mining tem a ver com Inteligência Artificial? Simples, Text Mining é composta por algumas áreas de conhecimentos e a Inteligência Artificial (Data Mining - Mineração de Dados) é uma delas. Além de IA, as outras principais áreas que contribuem com TM são: Processamento de Linguagem Natural, Estatística, Inteligência Computacional, Recuperação de Informação, Ciência Cognitiva, Mineração de Dados e Mineração na Web (do inglês, Web Minin (PUC-RIO, 2022).

Objetivos e Conceitos



Text Mining tem como objetivo identificar informações úteis e que não estão claras, mas que podem ser deduzidas. Além disso, por métodos tradicionais de consultas, essas informações não poderiam ser recuperadas. Isso porque os textos não possuem as informações de forma estruturada, o que torna necessário o uso de TM.

Com TM é possível recuperar textos representados em diversas formas, tais como: arquivos em formatos pdf, doc, txt, e-mails, web pages, textos em banco de dados, textos eletrônicos, entre outros.

Contudo, Text Mining pode ser encontrado a partir de outras terminologias, mas com o mesmo significado, tais como: Mineração de Dados em Textos (Text Data Mining) ou Descoberta de Conhecimento a partir de Bancos de Dados Textuais (Knowledge Discovery from Textual Databases), Busca de Informação (Information Seeking); Conhecimento Público não Descoberto (Undiscovered Public Knowledge); Recuperação de Conhecimento (Knowledge Retrieval). Essas últimas nomenclaturas são utilizadas devido ao fato de que o processo de TM pode ser feito a partir de técnicas de Knowledge Discovery in Databases – KDD ou Descoberta de Conhecimento em Bancos de Dados (MORAIS E AMBRÓSIO, 2007).

Seleção de Tipos de Informações

Para começar os processos de mineração de texto é necessário escolher (selecionar ou coletar) o texto ou a base de textos em linguagem natural, saber o que se pretende extrair e então definir o

que é esperado conseguir com a análise de texto, os obstáculos encontrados e assim analisar os resultados.

Por exemplo, no trabalho de MORAIS E AMBRÓSIO, 2007 "... o objetivo do processo de mineração de textos é extrair uma hierarquia de tópicos a partir de textos de um determinado domínio de conhecimento na área médica." A autora utiliza algoritmos de agrupamentos para auxiliar as tarefas de organização da informação textual, o resultado desses algoritmos prevê que um usuário interessado em um determinado documento deve estar interessado em outro documento desse grupo.

A seleção do texto ou textos é o primeiro item da Figura 1, que antecede as etapas de mineração e análise. Essas etapas serão mais detalhadas abaixo e são essas: definição do tipo de abordagem dos dados (análise semântica ou estatística), preparação dos dados, indexação e normalização, cálculo da relevância dos termos, seleção dos termos e pós-processamento (análise de resultados).



Pré-processamento: é nessa etapa (2) que as técnicas de Processamento de Linguagem Natural para estruturar os textos que serão analisados usadas. Segundo Filho, 2014 o "Processamento de Linguagem Natural (PLN) um conjunto de técnicas teórico-computacionais que visam representar dados textuais e processar a linguagem natural humana para diversas tarefas."

Indexação: na etapa 3, a indexação é feita. Os conceitos são extraídos do texto pela análise e trazidos em termos pelos padrões da indexação. É aqui que a identificação do documento e a definição dos pontos de acesso para consultas (GOMES, 2006).

Mineração: na etapa 4, é realizada aplicação de métodos e algoritmos para a identificação de padrões interessantes e extração de conhecimento.

Após a mineração, as informações coletadas podem ser classificadas em 2 tipos: Fato pode ser considerado o que já aconteceu e a opinião uma interpretação subjetiva, ou seja, interpretação dos fatos. A partir da necessidade de interpretar o subjetivo de uma maneira autônoma e posteriormente criar o conhecimento estruturado surgiu a Análise de Sentimentos (FILHO,2014).

Atividade Extra

Exemplo: Para se aprofundar no assunto desta aula assista o filme "Matrix" de Lily e Lana Wachowski.

Referência Bibliográfica

ARANHA, C. **Uma abordagem de pré-processamento automático para mineração de textos em português: sob o enfoque da inteligência computacional**. Departamento de Engenharia Elétrica, PUC-Rio. Rio de Janeiro. 2007.

FILHO, J. A. C. **Mineração de Textos: Análise de Sentimentos Utilizando Tweets Referentes à Copa do Mundo de 2014**. Universidade Federa. Ceará. Quixadá. 2014.

HESSLER, J. DE C. **Um método para apoio à leitura baseado no uso de uma ferramenta de mineração de texto.** Universidade Federal do Rio Grande do Sul. Porto Alegre. 2010.

MORAIS, E. A. M. e AMBRÓSIO, A. P. L. Mineração de Textos. Universidade Federal de Goiás. Goiás. 2007.

PUC-Rio. **Mineração de Textos: Fundamentos e Aplicações**. Rio de Janeiro. Disponível em <https://www.maxwell.vrac.puc-rio.br/11675/11675_3.PDF> Acesso em: 11 de Outubro de 2022

VALDATI, Aline de Brittos. **Inteligência Artificial – IA.** São Paulo: Contentus; 2020.

Ir para exercício